(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 August 2002 (15.08.2002)

PCT

(10) International Publication Number
WO 02/063460 A2

(51) International Patent Classification[7]: G06F 3/16

(21) International Application Number: PCT/GB02/00341

(22) International Filing Date: 25 January 2002 (25.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01103368.1 2 February 2001 (02.02.2001) CN

(71) Applicant *(for all designated States except US)*: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, NJ 10504 (US).

(71) Applicant *(for MG only)*: **IBM UNITED KINGDOM LIMITED** [GB/GB]; P.O. Box 41, North Harbour, Portsmouth, Hampshire PO6 3AU (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **YEH, James, Tien-Cheng** [US/CN]; No. 617, Purple Jade Villa, Chayang District, Beijing 100012 (CN). **SU, Hui** [CN/CN]; Room 101, Unit 1, Building 11, East Block, Tsinghua University, Beijing 100084 (CN). **WANG, Qianying** [CN/US]; No. 37, Angelll Ct., Apt. 314, Stanford, CA 94305 (US).

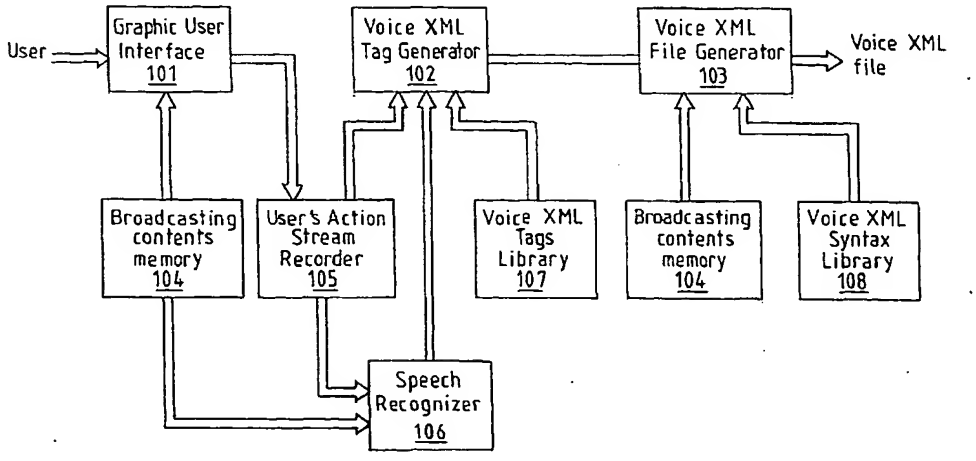
(74) Agent: **LING, Christopher, John**; IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Winchester, Hampshire SO21 2JN (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR AUTOMATICALLY CREATING VOICE XML FILE



(57) Abstract: This invention discloses a system for creating voice XML file automatically, comprising: a graphic user interface for defining a plurality of icons, wherein each of said icons corresponds to one or more attributes of voice XML; voice XML tag generator for interpreting said action stream based on a library of voice XML tags, generating the corresponding voice XML tags; and, voice XML file generator for combing the contents to be played with thte tags generated by the voice XML tag generator according voice XML syntax, for creating the voice XML file. This system can create the voice XML file for the TTS voice XML file or the real-time-recorded voice XML file.

WO 02/063460 A2

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND SYSTEM FOR AUTOMATICALLY CREATING VOICE XML FILE

The present invention generally relates to how to automatically create HTML(Hypertext Markup Language)file which can be used to broadcast message on the WWW (World wide Web) for network users, and more particularly to how to automatically create voice XML (voice XML) file which can be used to broadcast voice message on the WWW for network users.

Various browsers popularly used, such as the Netscape Navigator, have become one of the effective tools for network users to access the WWW. These browsers are textual and graphic user interfaces which aid network users in requesting and displaying information from the WWW. Besides text and graphics, information displayed by a browser may also include sound and hyperlinks and the like, thus the files displayed by a browser are often referred to as hypertext. If hypertext is used when conveying text information in a computer, not only is the linear construction of the information reserved, but also linking construction is added, The hypertext allows users jump-read text information, thereby facilitating users reading.

With the PvC devices becoming more and more popular, people are becoming unsatisfied with browsing network information only by way of reading, and audio broadcasting has become one of the major ways to browse network information for mobile users. However, browsing an audio file is not so easy as browsing a text file. The lack of interactive method is one of the main barriers. Under such a situation, user can only listen to broadcasted information passively. And there is no way for user to select information or find more detailed information when he/she listens to an interesting topic just as they are browsing the HTML files on network. Based on speech recognition technology, the technology to select information or find more detailed information based on dialog/conversation is being developed. voice XML is designed for this usage. However, it is not easy for an ordinary network user to write a voice XML file, which requires the user to have a good command of a large numbers of rules, syntax and definition of tags. Accordingly, the present invention provides a method and system for automatically creating voice XML file.

A method for automatically creating voice XML file in accordance with one aspect of the present invention comprises the steps of: providing a graphic user interface for defining a multiple of icons, wherein each of the icons corresponds to one or more attributes of voice XML; recording the action stream of users invoking the icons in the

graphic user interface; and interpreting the action stream based on a library of voice XML tags for creating the voice XML file.

A system for automatically creating voice XML file in accordance with another aspect of the present invention comprises: a graphic user interface for defining a multiple of icons, wherein each of the icons corresponds to one or more attributes of voice XML; voice XML tag generator for interpreting the action stream of a user based on a library of voice XML tags and generating the corresponding voice XML tags; and voice XML file generator 103 for combining the contents to be played with the tags generated by the voice XML tag generator according to voice XML syntax for creating the voice XML file.

These and other advantages and features of the present invention will become more clear from the description in conjunction with the accompanying drawings.

Fig. 1 shows how to add hyperlinks to a piece of audio and how user interacts with the hyperlinks;

Fig. 2 is a block diagram of a system for automatically creating voice XML file according to one preferred embodiment of the present invention;

Fig. 3 shows a graphic user interface according to one preferred embodiment of the present invention;

Fig. 4 shows a graphic user interface according to another preferred embodiment of the present invention;

Fig. 5 and Fig. 6 show an action stream for automatically creating voice XML file using the system shown in Fig. 2 according to one preferred embodiment of the present invention;

Fig. 7 is a flow chart showing the procedure of adding hyperlinks to TTS voice XML stream according to one preferred embodiment of the present invention;

Fig. 8 is a flow chart showing the procedure of adding hyperlinks to real-time-recorded audio voice XML stream according to one preferred embodiment of the present invention.

Fig. 1 describes how to add hyperlinks to a piece of audio and how user interacts with the hyperlinks. As shown in Fig. 1, for a sentence IBM is the biggest IT company in the world to be played, it can be formed as conventional TTS(text-to-speech)stream or real-time-recorded audio stream. To facilitate the user to obtain relevant detailed information on IBM, some attributes can be added, such as speech rendering attribute and linking attribute. As a result, when playing the above-mentioned sentence, audio browser can place emphasis on IBM in a different tone (or other attributes) so as to draw attention of the user. In the course of playing, when the user selects IBM in DTMF tone or in other ways, the audio browser retrieves the files at the address associated with the hyperlink and plays further information on IBM for the user. Thus, the user is not just listening to broadcasted information passively. When the user listens to an interesting topic, he/she may select information or obtain further information just as browsing the HTML files on the Internet. With the development of speech recognition technologies, user can not only select hypertext by DTMF tone, but also speak the hyperlink to be selected using the audio browser which has a barge-in voice recognition engine.

In spite of the advantages of voice XML as above mentioned, it is not easy for an ordinary network user to write voice XML file, which requires the user to have a good command of a large numbers of rules, syntax and definition of tags. Accordingly, the present invention provides a method and system for automatically creating voice XML file.

Fig. 2 is a block diagram of a system for automatically creating voice XML file according to one preferred embodiment of the present invention. As shown in Fig. 2, the system comprise: a graphic user interface 101 for defining a plurality of icons, wherein each of the icons corresponds to one or more attributes of voice XML; a voice XML tag generator 102 for interpreting the action stream based on a library of voice XML tags, generating the corresponding voice XML tags; and a voice XML file generator 103 for combining the contents to be played with the tags generated by the voice XML tag generator according to voice XML syntax for creating the voice XML file. According to one preferred embodiment of the present invention, the system may further comprise: a memory 104 for storing the contents to be played; a recorder 105 for recording the action stream of the user; speech recognizer 106; a voice XML tags library 107; a voice XML syntax library 108. When using the system to create voice XML file for a block of TTS stream, the user firstly interacts with the graphic user interface of the system. For a

block of TTS voice XML prompt, user can edit TTS stream in the editing area of the graphic user interface, marking or entering the parts needed to be added with the hyperlinks, and invoking the corresponding icons. Fig. 3 shows a graphic user interface according to one preferred embodiment of the present invention. The icons may correspond to one or more attributes of voice XML, such as:

Speech rendering attributes, including gender, tone and speed of the broadcaster, etc.;

Pointing functions realized by Barge-in functions; hyperlinks, etc.

The action stream recorder 105 of the system records users action stream, i.e., the procedure of users invoking the icons in the graphic user interface. Then, voice XML tag generator 102 interprets the action stream of the user based on voice XML tags library 107, generating corresponding voice XML tag. The voice XML file generator combines the contents to be played with the voice XML tags generated according to voice XML syntax so as to create the voice XML file.

When using the system to create voice XML for a block of real-time-recorded audio stream, a user also first interacts with the graphic user interface of the system. In the editing area of the graphic user interface, real-time-recorded audio stream is edited, parts to be added with the voice XML attributes are marked and entered, and the corresponding icons are called. For the real-time-recorded audio stream, when user enters the parts needed to be added with hyperlinks in the editing area, voice XML tag generator 102 of the system activates the speech recognizer 106 while interpreting users action stream for finding the parts that match the parts entered by the user in the real-time-recorded audio stream, so as to add voice XML attributes to the corresponding pars of the real-time-recorded audio stream. Examples in which the system automatically creates voice XML file for TTS stream and real-time-recorded audio stream are given below.

Example1:
< voice XML >

<prompt bargein=true><render.echo>IBM</render.echo>is the biggest IT company in the world</prompt>
<link next=http://www.ibm.com/vxml/mail.vxml>
<grammar>IBM</grammar>

```
<dtmf>1</dtmf>
    </link>

    </VXML>
```

Example 2:

```
</VXML>

<prompt bargein=true><audio src=ibmwelcome.wav></prompt>
<link next=http://www.ibm.com/vxml/mail.vxml>
<grammar>IBM</grammar>
<dtmf>1</dtmf>
    </link>

    </VXML>
```

In addition, when a user marks or enters the same parts needed to be added with the attributes of voice XML in the editing area of the graphic user interface for many times and the designated voice XML attributes are identical, or when a user marks or enters the parts needed to be added with the voice XML attributes in the editing area of the graphic user interface and has designated the attributes of voice XML, after the batch mode is selected, the voice XML file generator of the system processes all the stored TTS stream or all the real-time-recorded audio stream, adding the attributes of voice XML designated by the tag generator according to users invoking the icons to the parts that match the marked or entered parts needed to be added with the attributes of voice XML respectively, so that the efficiency of automatically creating voice XML file by the system will be improved greatly.

The above has described how to create voice XML file using the system shown in Fig. 2, briefly speaking, that is how to add various attributes of voice XML to TTS stream and real-time-recorded audio stream. In the various attributes of voice XML, one attribute is of significant importance: hyperlink. As above mentioned, if hypertext(hyperlink) is used when conveying text information in a computer, not only is the linear construction of the information reserved, but also a linking construction is added, which makes it possible for a reader to read text information in a jumping manner, thus facilitating users reading. Similarly, after hyperlinks are added to TTS stream or real-time-recorded audio stream, network user can select information or find more detailed information when he/she listens to voice XML files just as they are browsing the HTML

files. Therefore, based on the system for automatically creating voice XML file according to one preferred embodiment of the present invention, as shown in Fig. 4, a graphic user interface for adding hyperlink to voice XML file is provided in the graphic user interface. In the graphic user interface, the system automatically adds hyperlinks to TTS stream or real-time-recorded audio stream when a user marks or enters corresponding parts needed to be added with the hyperlinks and enters the corresponding hyperlink addresses.

Fig. 5 and Fig. 6 show the action stream that automatically creates voice XML file using the system as shown in Fig. 2 according to one preferred embodiment of the present invention. shown in Fig. 5, since the voice XML header has to be generated first, a user invokes the corresponding icon that matches the attributes of the header in the graphic user interface (such as the first icon from the left in Fig. 3). Then the user invokes the icon 302, and the system broadcasts the contents stored in the memory 104, for example, the main menu as follows, 0: weather, 1:stock, 2:ticket, 3:others. User enters the graphic user interface shown in Fig. 4, entering or marking 0: weather, and entering the linking address. Thereafter, following is to be done as indicated in Fig. 6. To begin with, similarly, user invokes the corresponding icons that match the attributes of the header in the graphic user interface, then TTS stream or real-time-recorded audio stream is broadcasted. When it comes to state or city, corresponding icons are invoked to add voice XML attributes (or voice XML hyperlinks)to them. After the user interacts with the system through the user interface in the above manner, the user action recorder records the whole operating procedure of the user, or more specifically, the procedure of invoking icons in the graphic user interface by the user. voice XML tag generator 102 interprets the action stream and generates the corresponding attributes of voice XML, and voice XML file generator 103 adds corresponding voice XML attributes to TTS stream or real-time-recorded audio stream so as to create the voice XML file.

Fig. 7 is a flow chart showing the procedure of adding hyperlinks to TTS voice XML stream according to one preferred embodiment of the present invention. As shown in Fig. 7, first the user edits TTS file in the editing area of the graphic user interface, as editing usually HTML files. Then the user enters or marks the parts needed to be added with voice XML hyperlinks, invokes corresponding icons, and enters corresponding hyperlink addresses.

Fig. 8 is a flow chart showing the procedure of adding hyperlinks to real-time-recorded audio voice XML prompt according to one preferred embodiment of the present invention, wherein when a user enters the parts needed to be added with voice XML hyperlinks in the editing area of graphic user interface, speech recognition technology has to be used to find in real-time-recorded audio stream the parts that match the parts to which voice XML hyperlinks need to be added.

The preferred embodiments have been described in conjunction with the accompanying drawings. It is understood by those skilled in the arts that various changes and modifications may be made without departing from the spirit and range of the invention. The invention encompasses all the changes and modifications, and the scope of the invention is only defined by the appended claims.

CLAIMS

1. A method for creating a voice XML file automatically, comprising:

providing a graphic user interface for defining a plurality of icons, each of said icons corresponds to one or more attributes of voice XML;

recording an action stream of a user invoking said icons in the graphic user interface; and

interpreting said action stream based on a library of voice XML tags for creating the voice XML file.

2. A method according to claim 1, characterized in that said graphic user interface comprises a graphic user interface for adding one or more audio hyperlinks for a voice XML file automatically, wherein each icon, defined in said graphic user interface, corresponds to a kind of hyperlinks.

3. A method according to claim 2, characterized in that when adding the hyperlinks for TTS voice XML file, the user edits the TTS voice XML file in the edit area of said graphic user interface, marks or enters the parts to be added with the hyperlinks, invokes the corresponding icons and enters the corresponding hyperlink addresses.

4. A method according to claim 2, characterized in that when the voice XML file for which the hyperlinks needs to be added is a real-time-recorded audio voice XML stream, the user edits the TTS voice XML file in the edit area of said graphic user interface, marks or enters the parts to be added with the hyperlinks, invokes the corresponding icons and enters the corresponding hyperlink addresses, and speech recognition technology is applied to find the parts in the real-time-recorded audio voice XML stream that match the parts entered by the user when interpreting said action stream based on a library of voice XML tags.

5. method according to claim 3 or claim 4, characterized in that when user marks or enters the same parts to be added with the hyperlinks in the edit area of the graphic user interface for many times and invokes the same hyperlink attributes, the hyperlinks for the whole TTS voice XML stream or the whole real-time-recorded audio voice XML stream are batch-added.

6. A system for creating voice XML file automatically, comprising:

a graphic user interface for defining a plurality of icons, wherein each of said icons corresponds to one or more attributes of voice XML;

a voice XML tag generator for interpreting said action stream based on a library of voice XML tags and generating the corresponding voice XML tags; and

a voice XML file generator for creating the voice XML file by combining the contents to be played with the tags generated by the voice XML tag generator according voice XML syntax.

7. A system according to claim 6, characterized in that said graphic user interface comprise a graphic user interface for adding audio hyperlinks for VioceXML file automatically, wherein each icon, defined in said graphic user interface, corresponds to a kind of hyperlinks.

8. A system according to claim 7, characterized in that when adding the hyperlinks for TTS voice XML stream, user edits the TTS voice XML file in the edit area of said graphic user interface, marking or typing the parts to be added the hyperlinks, invoking the corresponding icons and typing the corresponding hyperlink addresses.

9. A system according to claim 7, characterized in that when adding the hyperlinks for real-time-recorded audio voice XML stream, user edits the TTS voice XML file in the edit area of said graphic user interface, marking or typing the parts to be added the hyperlinks, invoking the corresponding icons and typing the corresponding hyperlink addresses, and when interpreting said action stream based on a library of voice XML tags, applying the speech recognition technology to find the parts in the real-time-recorded audio Voice XML stream that match the parts entered by the user.

10. A system according to claim 8 or claim 9, characterized in that when user marks or enters the same parts to be added the hyperlinks in the edit area of the graphic user interface for many times, and invoking the same hyperlinking attributes, adding batchly the hyperlinks for the whole TTS voice XML stream or the whole real-time-recorded audio voice XML stream.

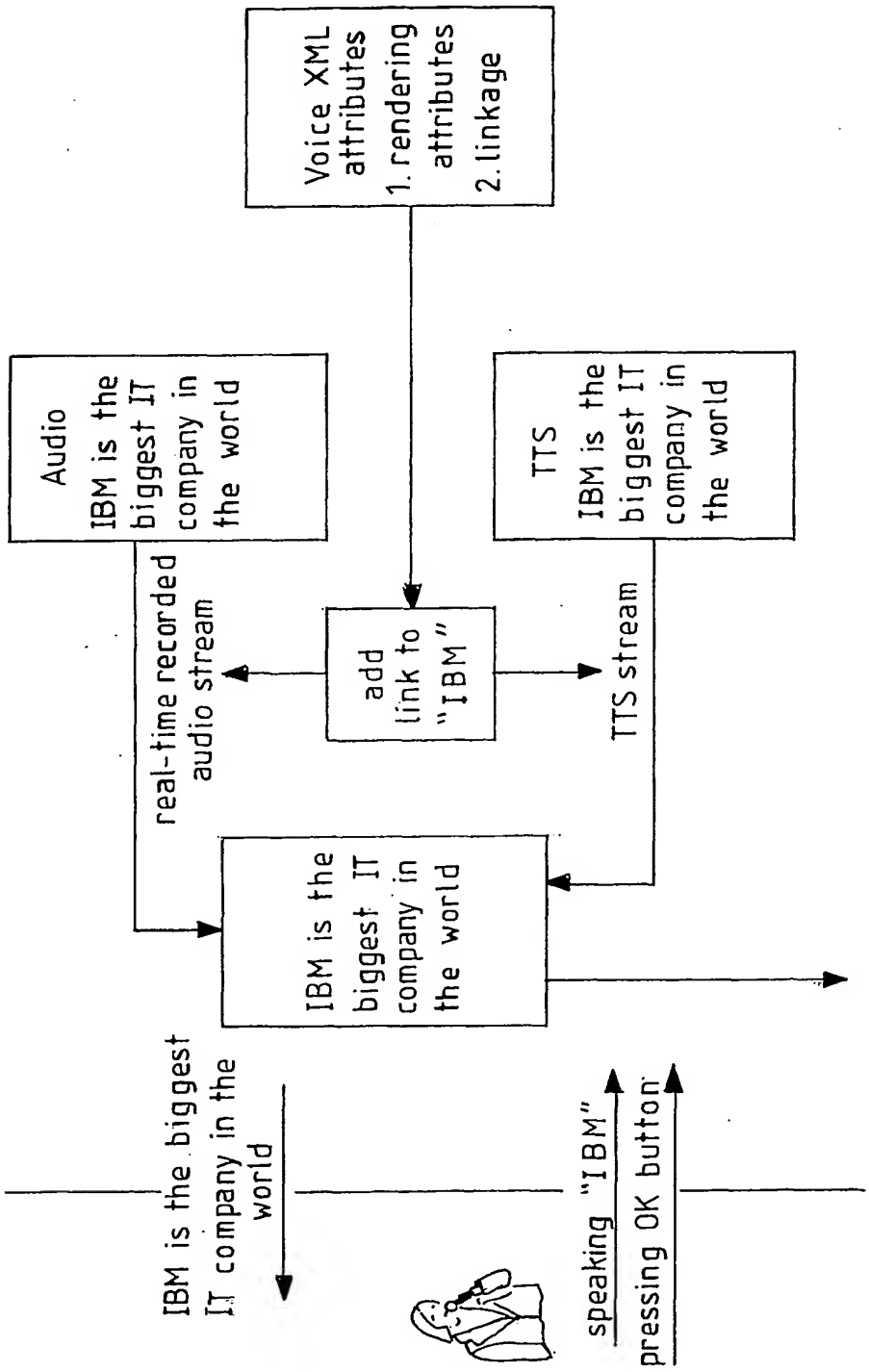

Voice XML attributes 1. rendering attributes 2. linkage
Audio IBM is the biggest IT company in the world
TTS IBM is the biggest IT company in the world
real-time recorded audio stream
add link to "IBM"
TTS stream
IBM is the biggest IT company in the world
IBM is the biggest IT company in the world
speaking "IBM"
pressing OK button

FIG. 1

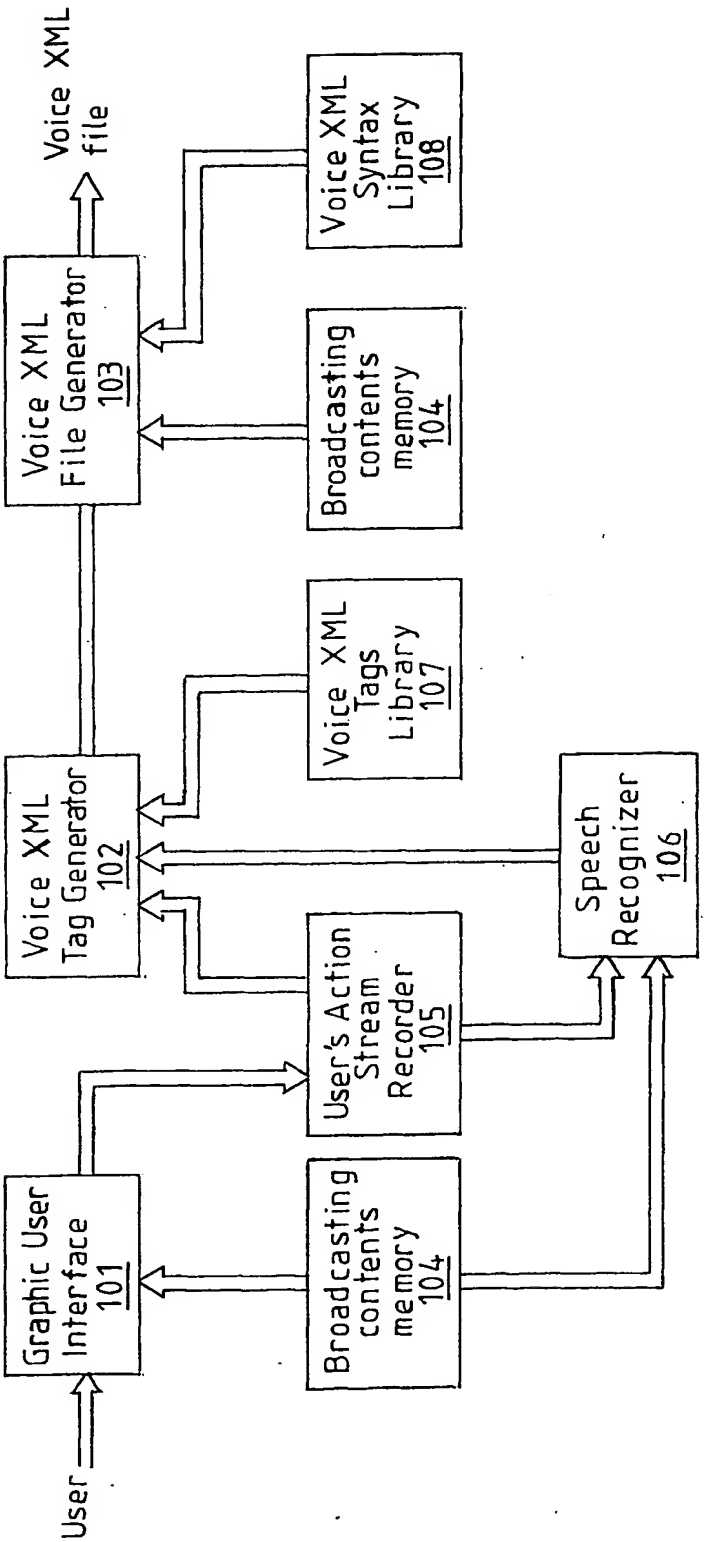

User
Graphic User Interface 101
Broadcasting contents memory 104
User's Action Stream Recorder 105
Voice XML Tag Generator 102
Voice XML Tags Library 107
Speech Recognizer 106
Voice XML File Generator 103
Broadcasting contents memory 104
Voice XML Syntax Library 108
Voice XML file

FIG. 2

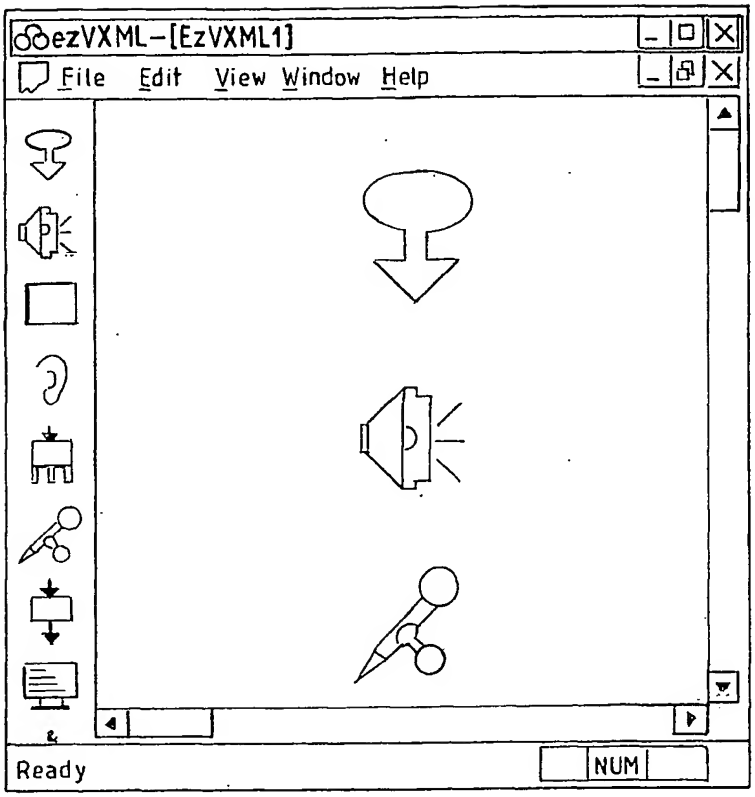


FIG. 3

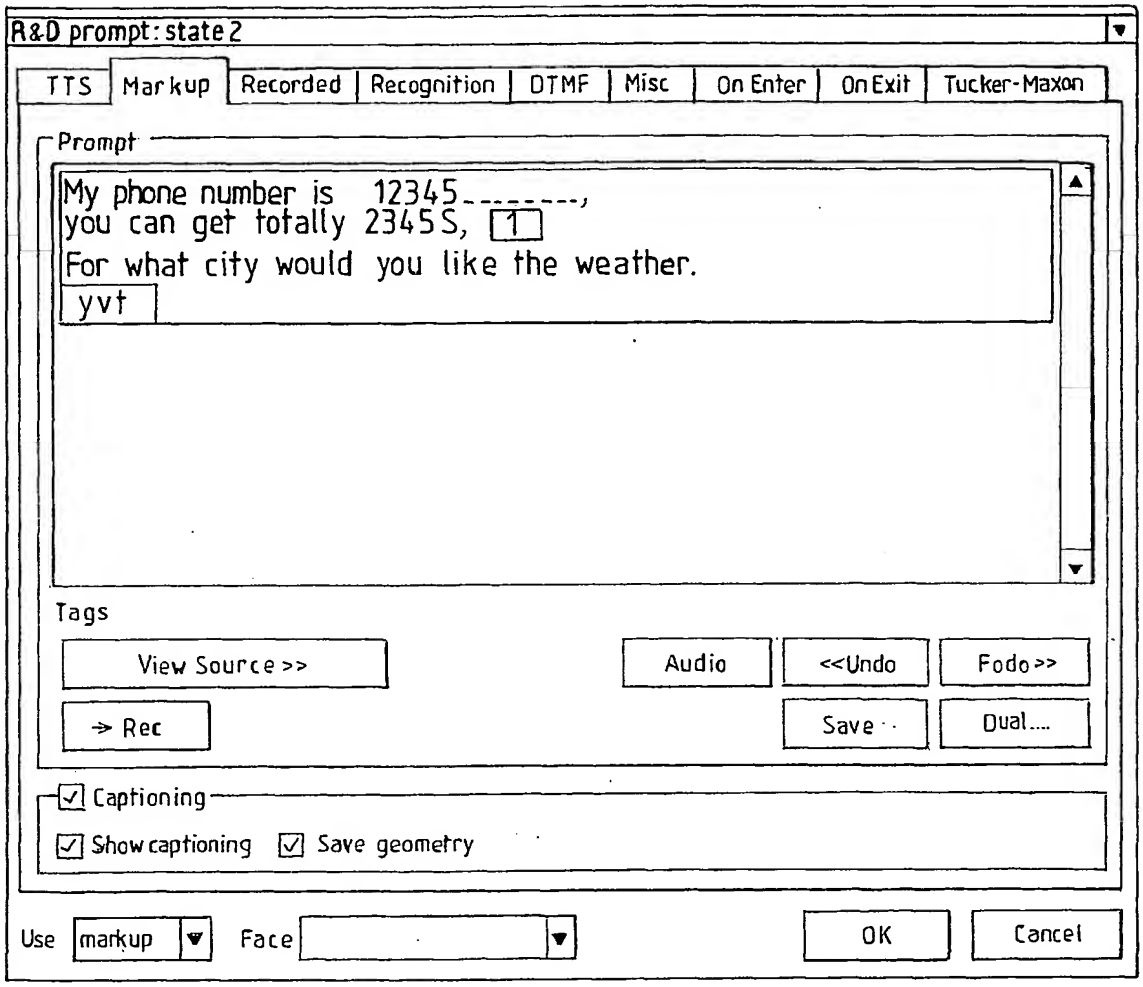

R&D prompt: state 2
TTS
Markup
Recorded
Recognition
DTMF
Misc
On Enter
On Exit
Tucker-Maxon
Prompt
My phone number is 12345_____,
you can get totally 2345S, 1
For what city would you like the weather.
yvt
Tags
View Source >>
Audio
<<Undo
Fodo >>
-> Rec
Save
Dual....
Captioning
Show captioning
Save geometry
Use
markup
Face
OK
Cancel

FIG. 4

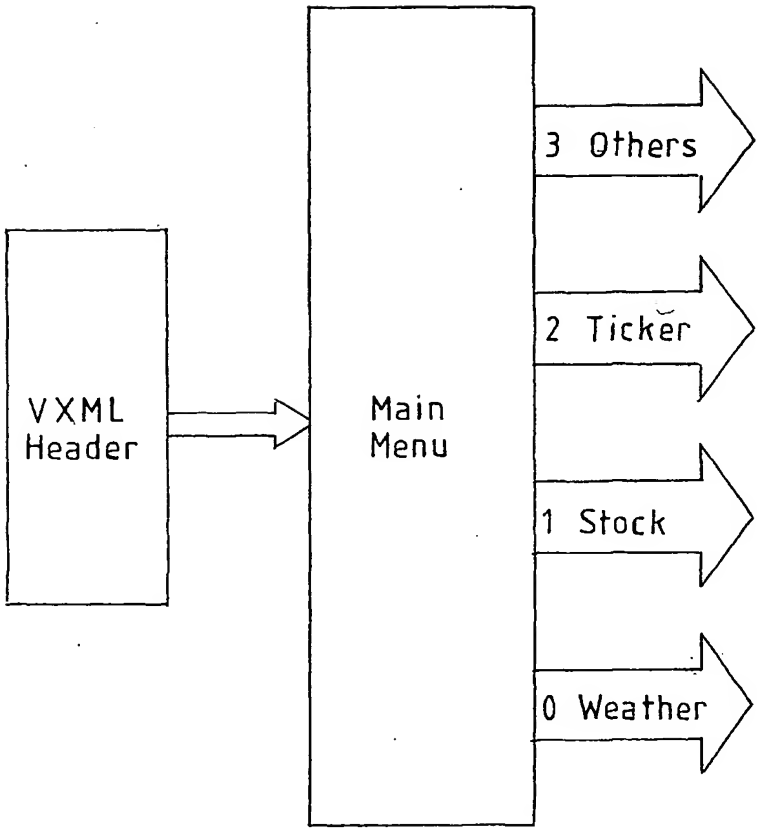


FIG. 5

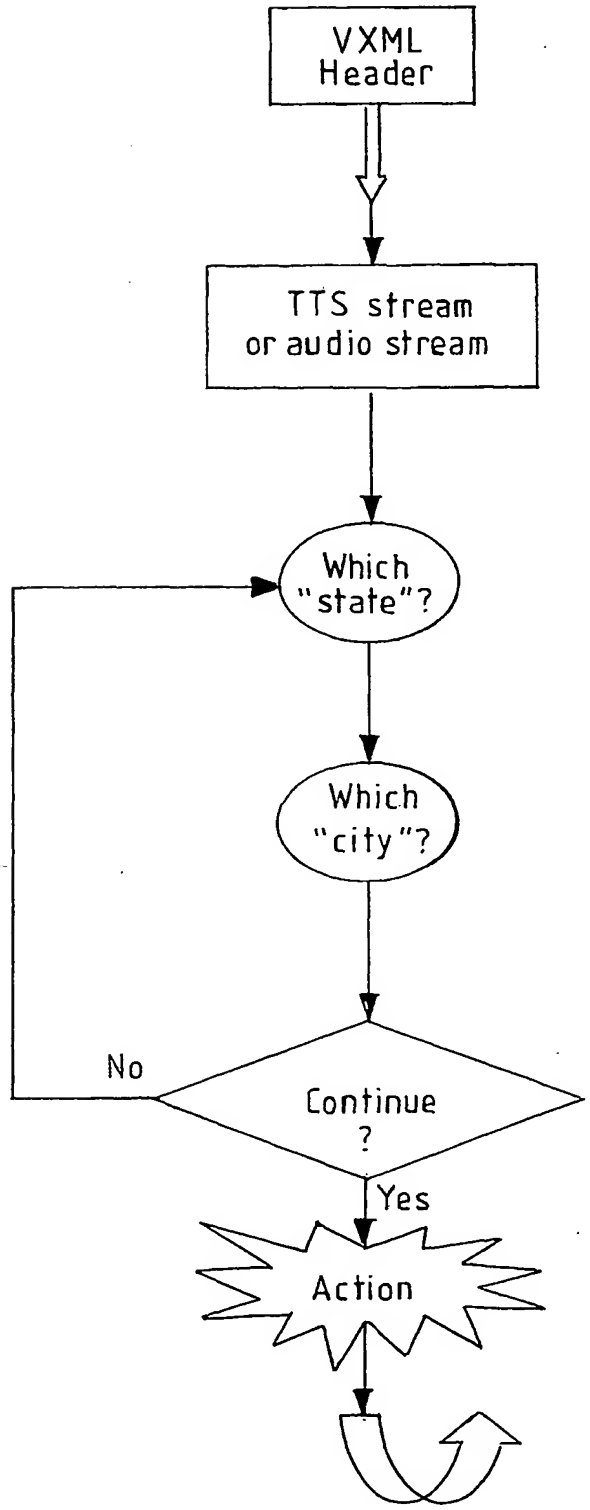

VXML Header
TTS stream or audio stream
Which "state"?
Which "city"?
No
Continue ?
Yes
Action

FIG. 6

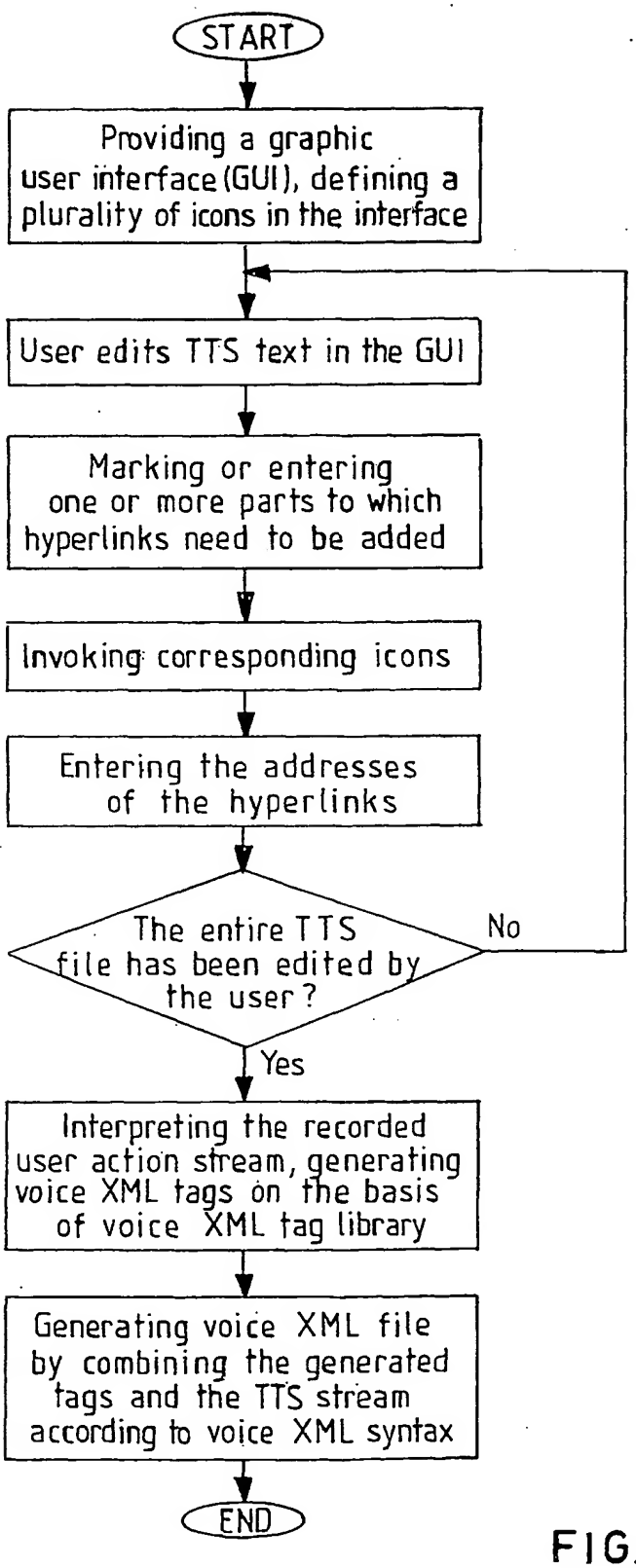

START
Providing a graphic user interface (GUI), defining a plurality of icons in the interface
User edits TTS text in the GUI
Marking or entering one or more parts to which hyperlinks need to be added
Invoking corresponding icons
Entering the addresses of the hyperlinks
The entire TTS file has been edited by the user?
No
Yes
Interpreting the recorded user action stream, generating voice XML tags on the basis of voice XML tag library
Generating voice XML file by combining the generated tags and the TTS stream according to voice XML syntax
END

FIG. 7

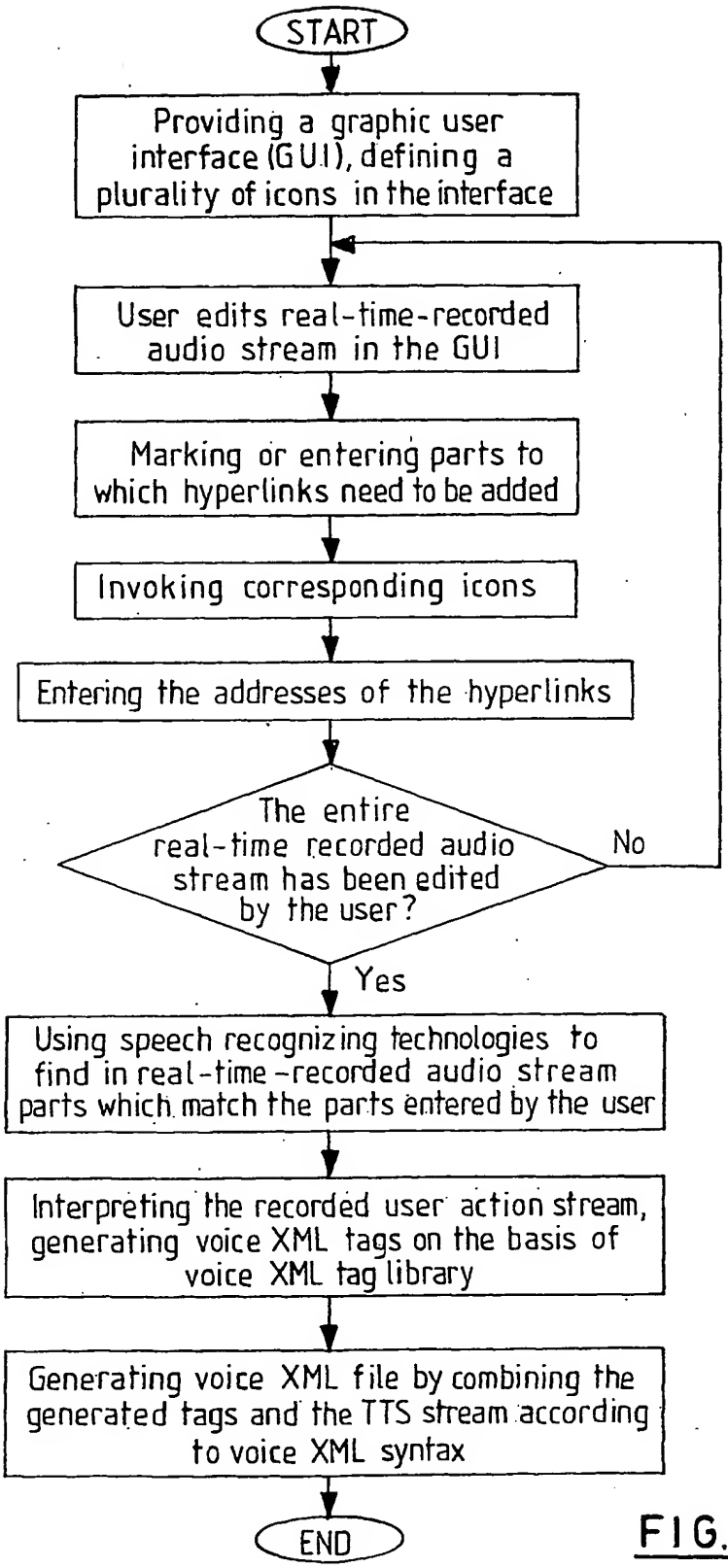

START
Providing a graphic user interface (GUI), defining a plurality of icons in the interface
User edits real-time-recorded audio stream in the GUI
Marking or entering parts to which hyperlinks need to be added
Invoking corresponding icons
Entering the addresses of the hyperlinks
The entire real-time recorded audio stream has been edited by the user?
No
Yes
Using speech recognizing technologies to find in real-time-recorded audio stream parts which match the parts entered by the user
Interpreting the recorded user action stream, generating voice XML tags on the basis of voice XML tag library
Generating voice XML file by combining the generated tags and the TTS stream according to voice XML syntax
END

FIG. 8